Network Detection and Association: Statistical Assessment of Performance

Kevin K. Anderson
Pacific Northwest Laboratory

Department of Energy Contract DE-AC06-76RLO 1830

Abstract

Statistical methods are needed to assess the performance of a global seismic monitoring system. This report considers the problems of estimating the detection capability and identifying misassociations. Detection capability is expressed empirically through a region- and magnitude-dependent detection probability function. Regional network information can substantially improve the estimation of the global network's detection capability. A quantitative measure of the strength of the association of an event is proposed. The proposed amplitude-based measure can be used to discriminate between real events and false associations. It quantifies the agreement among the observed and expected amplitudes of detecting and non-detecting stations.
Key words: network detection capability, association, network magnitude, goodness-of-fit.

Objective

A global seismic monitoring system is judged by its ability to detect and identify suspicious seismic events. Statistical methods are required to assess empirically the performance of all aspects of the monitoring system, from detection to discrimination. This report presents some preliminary research results in the areas of detection and association. Specifically, we report on the use of a catalog of seismic events from a regional network to help assess the detection capability of a global network in that region (Anderson 1995); and we propose a quantitative measure of the strength of event association as a way to identify both real events and false associations.

The global network region-dependent detection capability is expressed as a functional relation between the event magnitude and the probability of detection. Kelly and Lacoss (1969) and Ringdal (1986) have both taken this approach, and used the Gaussian probability distribution function as the basic functional form. We extend these authors' methods and show that event data from regional seismic networks can substantially improve the estimation of the region-dependent detection capability of the global network.

Automated association processes may result in large numbers of so-called false alarms, random phases that are associated to "create" an event where there really wasn't one. False alarms (spurious events) can be a problem for small events detected by relatively few stations. Quantitative measures of the strength of association can be constructed and used to identify false alarms. We propose one measure which is based on Ringdal's (1976, 1986) maximum-likelihood estimation of seismic magnitude. The measure quantifies the agreement among the observed and expected amplitudes of detecting and non-detecting stations.

Preliminary Research Results: Network Detection Probability

The empirically based Gutenberg-Richter magnitude-frequency relation (Gutenberg and Richter 1941) implies that the observed magnitudes of earthquakes in a particular region of homogeneous seismicity are randomly distributed according to a shifted exponential distribution. That is, the observed magnitudes m above some lower magnitude limit of interest m_{min} are well-modeled as a simple random sample from the probability density function

$$\beta \exp(-\beta(m - m_{min})), \qquad m > m_{min}. \tag{1}$$

The choice of m_{min} is important. For the problem of determining the value of β, a value of m_{min} is chosen, such that network detection is complete above it (for example, see Bender 1983). For the problem of estimating the probability of detection as a function of observed magnitude, a value of m_{min} below the magnitude of complete detection is chosen such that the Gutenberg-Richter magnitude-frequency relation is assumed to hold. This m_{min} value is usually greater than or equal to the magnitude of the smallest event observed by the network. The assumption that Equation (1) holds is critical because it serves as "ground-truth" in the estimation of detection probabilities. Non-Gutenberg-Richter frequency-magnitude relationships have been observed (Taylor et al. 1990).

Kelly and Lacoss (1969) derived a maximum likelihood estimation method for the simultaneous estimation of the parameters of the Gutenberg-Richter magnitude-frequency relation and the parameters of a probit model for the probability of detection using continuous magnitude data; that is, they modeled the network detection probabilities $P(m)$ as a function of magnitude m using the Gaussian distribution function

$$P(m) = \Phi((m-c)/d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(m-c)/d} \exp(-y^2/2)\, dy.$$

They derived their method assuming that, when grouped into non-overlapping magnitude intervals, the number of earthquakes occurring in each interval are independent Poisson random variables. Here, we extend the Kelly and Lacoss (1969) method to provide a joint analysis of two seismic networks.

We assume that two catalogs of seismic events are available covering the same temporal and spatial area; one from a regional network (A) in the region of interest and the other from a large globally distributed network (B). We first consider the case where the two networks' detections in the region of interest are

statistically independent. We assume that event matching between the two catalogs is accurate. We assume that the data available for analysis is a breakdown by magnitude grouping of the number of events detected uniquely or jointly. Y_{ij} is defined as the number of magnitude m_i events of type j, where $j = 1,2,3$ corresponds to events detected by both networks, by network A alone, or by network B alone, respectively; and the magnitudes range from m_{min} to m_{max}. Given that the number of earthquakes N has a Poisson distribution with mean λ, the Y_{ij} are independent Poisson random variables with means

$$\exp(\alpha - \beta m_i) p_{ij},$$

where

$$p_{i1} = \Phi((m_i - c_A)/d_A)\Phi((m_i - c_B)/d_B),$$
$$p_{i2} = \Phi((m_i - c_A)/d_A)(1 - \Phi((m_i - c_B)/d_B)),$$
$$p_{i3} = (1 - \Phi((m_i - c_A)/d_A))\Phi((m_i - c_B)/d_B).$$

The parameter α depends on λ and β through

$$\exp(\alpha) = \lambda / \sum_{j=1}^{n} \exp(-\beta m_j).$$

Each network has its own detection probability parameters subscripted by A and B. The parameter c is the 50% detection magnitude and the parameter d affects the steepness of the probability-of-detection curve.

Estimation of the six parameters and their uncertainties can be accomplished by maximum likelihood or by using a two-stage nonlinear least squares approach. A goodness-of-fit test to judge the adequacy of the model can then be performed. Let "hats" indicate parameter estimates, $\hat{Y}_{ij} = \exp(\hat{\alpha} - \hat{\beta} m_i)\hat{p}_{ij}$, and

$$X^2_{GOF} = \sum_{i=1}^{n} \sum_{j=1}^{3} (Y_{ij} - \hat{Y}_{ij})^2 / \hat{Y}_{ij}.$$

To test the adequacy of the model, X^2_{GOF} is compared with a chi-square distribution with $(3n - 6)$ degrees of freedom. We estimate $p \times 100\%$ detection thresholds for the two networks as

$$\hat{c}_A + \hat{d}_A \Phi^{-1}(p) \qquad \text{and} \qquad \hat{c}_B + \hat{d}_B \Phi^{-1}(p),$$

where $\Phi^{-1}(\cdot)$ is the inverse of the Gaussian distribution function.

With a simple modification, we can account for one type of dependent networks. Suppose the networks are such that if an event is detected by the global network (B), it is also detected by the regional network (A). In this case, there are no type 3 events, where network B detected, but network A did not; that is, all $Y_{i3} = 0, i = 1, 2, \ldots, n$. This type of dependence is easily accounted for by redefining the p_{ij} as follows:

$$p_{i1} = \Phi((m_i - c_B)/d_B),$$

$$p_{i2} = \Phi((m_i - c_A)/d_A)(1 - \Phi((m_i - c_B)/d_B)),$$

and $p_{i3} = 0$. Using the Poisson model, maximum likelihood estimation or two-stage nonlinear least squares estimation follows as above. Further extensions or modifications, such as replacing the probit probability-of-detection models with other models, are possible.

To demonstrate the method developed above, we considered event catalogs from the Pacific Northwest Seismograph Network (PNSN) and the U. S. Geological Survey (USGS) for 1990-1992 (Anderson 1995). The PNSN has more than 100 stations in the Pacific Northwest.

The three-year PNSN data set analyzed by Anderson (1995) contained 3032 Washington state earthquakes, of which the USGS network detected 123. We determined earthquake matches using location and origin time. For most of the 123 earthquakes, the USGS catalog did not report a network m_b magnitude, so

the PNSN P-coda magnitudes were used in the analysis. The magnitude distribution of these jointly detected events is displayed in Figure 1. No earthquakes were detected by only the USGS. The two networks are not independent, because the USGS receives data from eight PNSN stations by telephone lines in real time (and, in fact, the locations for most of these 123 USGS catalog events are the PNSN locations). Therefore, we analyzed the data using the dependent two-network model.

We had to remove one event from the analysis, a magnitude 3.2 event that was not listed in the USGS catalog. This single "missing" event caused the goodness-of-fit test to reject the model; that earthquake should have been detected by USGS with high probability. Table 1 contains the results of a two-stage nonlinear least squares analysis, with that event ignored. The estimates of the 90% detection thresholds for PNSN and USGS for Washington state are 2.562 and 2.844, with standard errors of 0.224 and 0.028, respectively.

Some caveats must accompany the 90% detection thresholds given above. The analysis was performed to illustrate the statistical method. While one 90% detection threshold might adequately describe the detection capability of the global network in that region, the detection capability of the PNSN regional network is known to be much more spatially heterogeneous. The one 90% detection threshold given for the PNSN is a rough spatial average. Further, the USGS processing is not well-understood; the very sharp cutoff at magnitude 2.5 may merely reflect a procedural decision to exclude events rather than a physical detection limitation.

Table 1: Dependent two-network analysis results: parameter estimates and standard errors

Method	$\hat{\alpha}$	$\hat{\beta}$	$\hat{c}_{PNSN}$	$\hat{d}_{PNSN}$	$\hat{c}_{USGS}$	$\hat{d}_{USGS}$
Two-Stage NLS	9.680	2.407	1.750	0.634	2.607	0.185
	(0.611)	(0.211)	(0.189)	(0.031)	(0.022)	(0.006)

Preliminary Research Results: Measuring Strength of Association

Determining misassociations for a seismic network is challenging. There are two basic types of errors in association. Events may be missing from a network seismicity catalog because the waveforms were not associated or the catalog may contain spurious events, those events which were misassociated based on signals from multiple events. The first type of error, missed association, is different from non-detection. It represents a failure of the association process when there were enough detections to make an association. Events which were missed associations can be discovered through comparison with seismicity catalogs from other networks, or, in the case of explosions, through a ground-truth source. The second type of error, spurious events, may be detected through visualization methods which allow the analyst access to all the information which made up the events. The analyst then bases decisions on his prior experience; he "knows a bad event when he sees one." Quantitative methods of measuring the strength of association are required to reduce the amount of work the analyst must do "by eye." These methods could be used to remove weakly associated events from final seismicity catalogs with or without analyst intervention.

When a spurious event is constructed from the waveforms of multiple events, it is usually the case that the arrival times are consistent, that is, they agree with the expected arrival times from an event placed at the location of the spurious event. By the somewhat random nature of such misassociations, it is expected that other parameters of interest, such as amplitudes and azimuths, would be less consistent, and misassociations could be identified by the larger-than-expected variability among the data from different detecting stations. In less extreme cases, with only a few of the waveforms associated in error, the effect might only be a slight increase in the size of the location error ellipses, or increased uncertainty in network parameter estimates. A goodness-of-fit statistic for the network maximum likelihood magnitude estimate is developed here.

Ringdal (1976) presented a maximum likelihood approach to the estimation of network m_b values. The approach successfully removes the network magnitude bias problem that other studies identified (e.g., Herrin

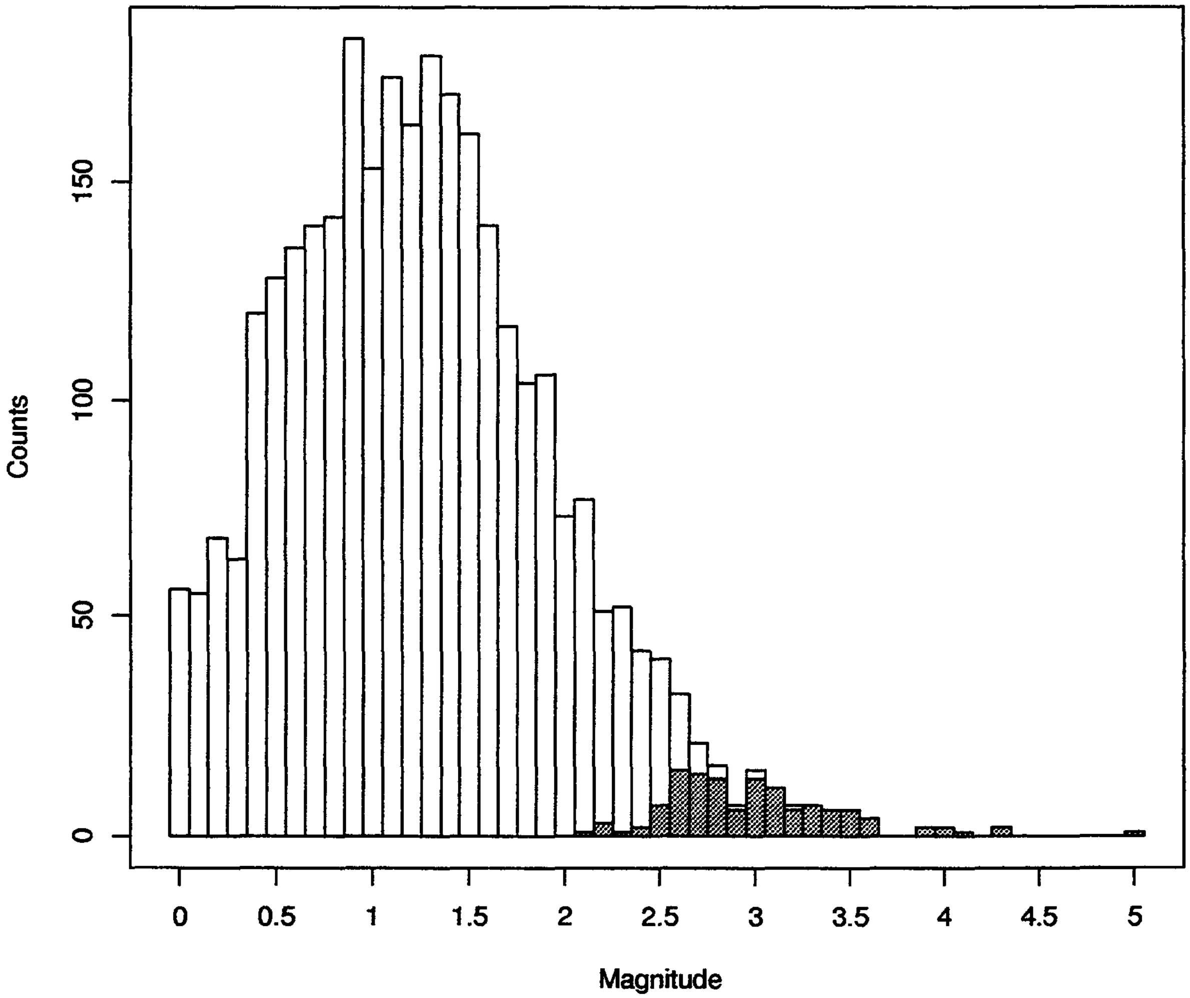

Figure 1: Frequency of State of Washington events, 1990-92, detected by the PNSN and the USGS seismic networks. The unshaded bars represent the number of earthquakes detected by only the PNSN. The shaded bars represent the number of earthquakes detected by both networks.

and Tucker (1972)), by including the information from non-detecting stations. We extend that approach to assess the strength of the association of a seismic event. Since a spurious event may have increased variability in the distance-corrected amplitudes, and may also have no detections at stations that would have detected a real event, we develop a goodness-of-fit statistic that is sensitive to deviations from expected amplitudes and detections.

We assume Ringdal's (1986) model, that for a seismic network of N stations, $g_i, i = 1, 2, \ldots, N$, denotes the P-wave detection thresholds of the stations and $y_i, i = 1, 2, \ldots, N$, denotes the P-wave signals of a seismic event of true bodywave magnitude μ.. We assume that g_i and y_i are Gaussian distributed as follows

$$g_i \sim N(G_i, \gamma_i^2)$$

$$y_i \sim N(\mu - Q_i + B_i, \sigma_i^2),$$

where G_i is the average station detection threshold, Q_i is a distance-depth correction factor (dependent on the locations of the event and the station), B_i is the average station magnitude bias, and γ_i and σ_i are the standard deviations.

Ringdal (1986) includes in the network magnitude estimation only those stations within the teleseismic range of 21° to 100° from the event. These stations fall into four groups: Group A, stations reporting a P-wave detection and an associated signal level; Group B, stations reporting a P-wave detection with no associated signal level; Group C, stations operable but not reporting a P-wave detection; and Group D, stations which are inoperable or off-line. The Group D stations provide no information and can be omitted from the analysis. We could have defined a fifth group, stations outside the range of 21° to 100° from the event, and subsequently omitted them from the analysis.

The likelihood function $L(\mu)$ is essentially that described by Ringdal (1986):

$$L(\mu) = \prod_{i \in A} f_i(\mu) \prod_{i \in B} h_i(\mu) \prod_{i \in C} (1 - h_i(\mu)),$$

where

$$f_i(\mu) = \frac{1}{\sigma_i} \phi((\mu - y_i - Q_i + B_i)/\sigma_i)$$

$$h_i(\mu) = \Phi\left(\frac{\mu - G_i - Q_i + B_i}{\sqrt{\sigma_i^2 + \gamma_i^2}}\right)$$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(x^2/2)$$

and $\Phi(\cdot)$ is the Gaussian distribution function defined earlier. All parameters but μ are known. The network maximum likelihood estimate $\hat{\mu}$ of the event magnitude μ is obtained by numerically maximizing $L(\mu)$. The usual estimate of the variance of $\hat{\mu}$ is the reciprocal of the second derivative of the negative-log-likelihood function evaluated at $\hat{\mu}$.

A goodness-of-fit statistic for the network maximum likelihood magnitude $GOF(m_b)$ is

$$GOF(m_b) = -2\log(L(\hat{\mu})) - \sum_{i \in A} \log(2\pi\sigma_i^2).$$

This goodness-of-fit statistic is approximately distributed as a chi-square random variable with $k = n - \Delta_{GOF} - 1$ degrees of freedom, where n is the number of stations within 21° to 100° of the event and

$$\Delta_{GOF} = \sum_{i \in B} \chi(h_i(\hat{\mu}) - 0.97) + \sum_{i \in C} \chi(0.03 - h_i(\hat{\mu}))$$

and $\chi(\cdot)$ is the indicator function; if $x \geq 0$, then $\chi(x) = 1$, else $\chi(x) = 0$. The adjustment Δ_{GOF} reduces the degrees of freedom for the stations that provide no information relevant to the magnitude. A large value

of $GOF(m_b)$ is evidence that the event is a potential spurious event. Excessive variability in the P-wave amplitudes or missing detections will increase the value of $GOF(m_b)$. An approximate statistical test of the null hypothesis that an event is real is to reject if $GOF(m_b)$ exceeds the appropriate percentile of the chi-square distribution with k degrees of freedom. Using the 95th percentile as the cutoff sets the probability of falsely classifying a true event as spurious at approximately 5 %.

Power studies can be performed to determine the ability of the method to accurately identify spurious events. The power of a statistical test is the probability of rejecting the null hypothesis when it is false. In this case, the power is the probability that the goodness-of-fit statistic $GOF(m_b)$ will indicate that a spurious event is indeed spurious. The power of this test for a particular global network depends on the magnitude and hypocenter of the event, and on how one models a spurious event.

As an example of estimating the power, we considered the globally distributed network of 115 stations described in Ringdal (1986). We modeled a spurious event with a nominal m_b of 3.5 and located at 17°S and 170°E, under combinations of two factors; the first, that the standard deviations of the amplitudes are multiplied by an inflation factor in the range of 1 to 2.5 and second, that as many as two "good" stations did not detect the event, where "good" stations are those that would be expected to detect the event and to provide amplitudes for the the computation of network magnitude. We set the cutoff at the 95th percentile of the chi-square distribution. Further details of the Monte Carlo analysis are omitted. The results from the Monte Carlo analysis are used to generate the three curves in Figure 2. They give the probability of classifying the spurious event as spurious for the range of standard deviation inflation factors and zero, one, or two missing good stations.

Interpretation of the curves is as follows. If all of the stations detect the event as expected and the signal-to-noise ratios are as expected, then the event "looks" real as far as amplitudes are concerned and the probability of classifying the event as spurious is just 5%, the same as for a real event. If all of the stations detect the event as expected and the signal-to-noise ratios are about 1.5 times as expected, then the probability of classifying the event as spurious is about 40%. If two good stations miss the event and the signal-to-noise ratios are about 1.5 times as expected, then the probability of classifying the event as spurious is about 80%. These results indicate the goodness-of-fit statistic $GOF(m_b)$ can provide reasonable identification of spurious events.

Recommendations and Future Plans

The method of Kelly and Lacoss (1969) has been used to determine the region-dependent global network event detection capabilities. We extended the method to include events detected by a regional network. The joint modeling of the detections of the two networks can provide better estimates of the global network's detection probabilities, especially if the number of events detected by the global network is small. The global network's detection capabilities can be tracked in time. The problems associated with matching up events in the two catalogs were not considered here, but may complicate the two-network analysis. The magnitude used in analysis can be the global network magnitude, the regional network magnitude, or some combination of the two.

A goodness-of-fit statistic based on the detection pattern and the amplitudes of the detecting stations was shown to be a reasonable quantitative measure of the strength of event association. Combined with current visualization techniques, the $GOF(m_b)$ statistic can reliably identify misassociated events. Future research will consider including amplitudes and detections from stations at regional distances, within 21° from the event. We will also consider goodness-of-fit statistics based on other seismic parameters, such as azimuth and first motion. The combining of various goodness-of-fit statistics is also an issue. Analysis of catalogs of events from regional networks can also provide information about missed events and spurious events.

Acknowledgement. The author thanks Alan Rohay, Don Daly, Rich Hanlen, and Wes Nicholson for useful discussions. The research was done at the Pacific Northwest Laboratory, a multiprogram laboratory operated for the U.S. Department of Energy under Contract DE-AC06-76RLO 1830.

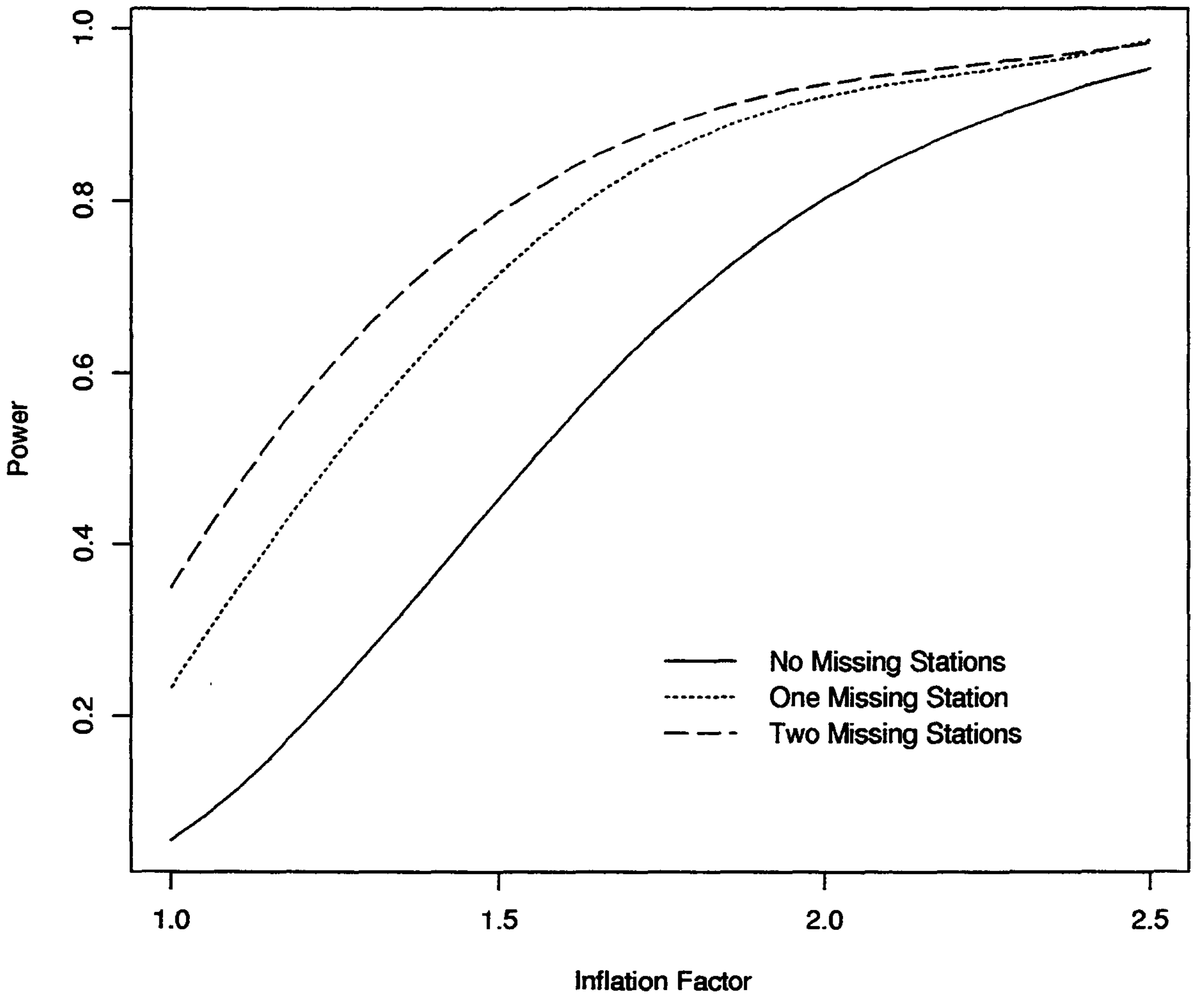

Figure 2: Probability of Identifying Spurious Events.

References

[1] Anderson, K. K. (1995). On the estimation of network detection probabilities, *Technical Report PNL-SA-264180,* Pacific Northwest Laboratory, Richland, Washington.

[2] Bender, B. (1983). Maximum likelihood estimation of b values for magnitude grouped data, *Bull. Seism. Soc. Am.* **73**, 831-851.

[3] Gutenberg, B. and C. F. Richter (1941). Seismicity of the Earth, *Geol. Soc. Am., Spec. Pap.* **34**, 1-133.

[4] Herrin, E., and W. Tucker (1972). On the estimation of bodywave magnitude, *Technical Report to AFOSR,* Dallas Geophysics Laboratory, Southern Methodist University, Dallas, Texas.

[5] Kelly E. J. and R. T. Lacoss (1969). Estimation of seismicity and network detection capability, Technical Note 1969-41, Lincoln Laboratory, MIT, Lexington, Massachusetts.

[6] Ringdal, F. (1976). Maximum-likelihood estimation of seismic magnitude, *Bull. Seism. Soc. Am.* **66**, 789-802.

[7] Ringdal, F. (1986). Study of magnitudes, seismicity, and earthquake detectability using a global network, *Bull. Seism. Soc. Am.* **76**, 1641-1659.

[8] Taylor, D. W. A., J. A. Snoke, I. S. Sacks, and T. Takanami (1990). Nonlinear frequency-magnitude relationships for the Hokkaido Corner, Japan, *Bull. Seism. Soc. Am.* **80**, 340-353.